Sparse Representations for Structured Noise Filtering

Bishnu P. Lamichhane* and Laura Rebollo-Neira*

February 1, 2008

Abstract

The role of sparse representations in the context of structured noise filtering is discussed. A strategy, especially conceived so as to address problems of an ill posed nature, is presented. The proposed approach revises and extends the Oblique Matching Pursuit technique. It is shown that, by working with an orthogonal projection of the signal to be filtered, it is possible to apply orthogonal matching pursuit like strategies in order to accomplish the required signal discrimination.

1 Introduction

The problem of structured noise filtering is introduced in [1], where a number of relevant signal processing applications are discussed. It can be posed as follows: consider that a signal f, represented as an element of an inner product space $\mathcal{H}$, is produced by the superposition of two components, f_1 and f_2, each of which belongs to a different subspace of $\mathcal{H}$. More precisely, $f = f_1 + f_2$ with $f_1 \in \mathcal{S}_1 \subset \mathcal{H}$ and $f_2 \in \mathcal{S}_2 \subset \mathcal{H}$. Structured noise filtering (to be also termed signal discrimination or signal splitting) consists of singling out a particular component from the signal f. Provided that $\mathcal{S}_1$ and $\mathcal{S}_2$ are given, such that $\mathcal{S}_1 \cap \mathcal{S}_2 = \{0\}$, one component, say f_1, can be extracted from f by an oblique projection onto $\mathcal{S}_1$ and along $\mathcal{S}_2$. On the contrary, the situation $\mathcal{S}_1 \cap \mathcal{S}_2 \neq \{0\}$ implies that the signal decomposition is not unique and the splitting can not be tackled in a straightforward manner by oblique projections. Moreover, even when theoretically the condition $\mathcal{S}_1 \cap \mathcal{S}_2 = \{0\}$ is satisfied, if the subspaces $\mathcal{S}_1$ and $\mathcal{S}_2$ are not well separated, the construction of the corresponding projector becomes ill posed. Consequently, the signal splitting can not be achieved by numerical calculations in finite precision arithmetics. Here we focus on such a situation. *We assume that the given subspaces $\mathcal{S}_1$ and $\mathcal{S}_2$ are theoretically disjoint, but close enough to yield an ill posed problem.*

Our proposal for the numerical realization of the signal splitting is focussed on the search of a subspace of the given $\mathcal{S}_1$, where a class of signals is considered to lie. It will be assumed throughout the paper that the class of signals to be considered is K-sparse in a spanning set for $\mathcal{S}_1$. By this we mean that given a spanning set for $\mathcal{S}_1$ the corresponding linear superposition of a signal has at most K nonzero coefficients. The K-value should be less than or equal to the dimension of the subspace $\mathcal{S}_r \subset \mathcal{S}_1$ for which the construction of an oblique projection onto itself, and along $\mathcal{S}_2$, is well conditioned. This assumption is quite realistic, considering that in practice there is often a lack of complete knowledge on the actual subspace $\mathcal{S}_1$ and to be on the safe side one may overestimate it.

*Aston University, Aston Triangle, B4 7ET, UK, b.p.lamichhane@aston.ac.uk, rebollol@aston.ac.uk

The main motivation of this paper is to highlight the essential role that sparse representations play in the problem of structured noise filtering. Such representations have been the subject of considerable work over the last ten years [2–14]. We will dedicate special attention to discuss and illustrate 'why' and 'how' sparse representations are relevant in the present context.

A technique, termed Oblique Matching Pursuit (OBMP), has been recently advanced in relation to the above described problem [15]. Such a technique evolves by stepwise selection of the sought subspace. The selection criterion is based on the consistency principle [16, 17]. In this communication we revise and extend the OBMP technique. We show that by working with a particular projection of the signal at hand, rather than with the signal itself, one can make use of previously proposed orthogonal matching pursuit like methodologies, so as to look for the signal subspace yielding the correct splitting.

The paper is organized as follows: In Section 2 we introduce the mathematical setting for signal representation to be adopted here, together with a discussion on the construction of oblique projections. Section 3 highlights the importance of the search for sparse solutions in the construction of oblique projectors for structured noise filtering. The proposed strategy is discussed in Section 4. The conclusions are presented in Section 5.

2   Mathematical Framework

We consider a signal, f, to be an element of an inner product space $\mathcal{H}$. The square norm $||f||^2$ is then induced by the inner product that we indicate as $\langle f, f \rangle$ and is defined in such a way that if a is a number, $\langle af, f \rangle = a^* \langle f, f \rangle$, with a^* representing the complex conjugate of a. For the purpose of this contribution we assume that all the signals of interest belong to some finite dimensional subspace $\mathcal{V}$ of $\mathcal{H}$. Thus, there exists a finite set $\{v_i \in \mathcal{H}\}_{i=1}^M$ spanning $\mathcal{V}$. Consequently, for every signal in $\mathcal{V}$ there is a set of numbers $\{c_i\}_{i=1}^M$ which allows us to express the signal as the linear superposition

$$f = \sum_{i=1}^{M} c_i v_i,$$

which is also called *atomic decomposition*.

Although a signal was defined as an element of an abstract inner product space, for processing purposes we need a numerical representation of such an object. The process of transforming a signal into a number is refereed to as *measurement* or *sampling*. The mathematical operation performing such a transformation is then a *functional*. Since considerations will be restricted to linear measurements, we represent them by *linear functionals*. Thus, making use of Riesz theorem [18] we can express a linear measurement as $m = \langle w, f \rangle$ for some $w \in \mathcal{H}$. Considering now M measurements m_i, $i = 1, \ldots, M$, each of which is obtained by a measurement vector w_i, we have a numerical representation of f as given by

$$m_i = \langle w_i, f \rangle, \quad i = 1, \ldots, M. \tag{1}$$

The question concerning the possibility of reconstructing $f \in \mathcal{V}$ from measurements obtained with vectors in a different subspace has been addressed in [16, 17, 19, 20]. It is in principle obvious that every signal in $\mathcal{V}$ can be reconstructed from vectors $\{w_i \in \mathcal{H}\}_{i=1}^M$ spanning a subspace $\mathcal{W} \subset \mathcal{H}$, provided that those vectors give rise to a representation of any projector onto $\mathcal{V}$. The difference in using one projector or another appears when the projector acts on signals outside a subspace. We summarize next some features relevant to the construction of projectors.

2.1 Oblique projectors

Every idempotent operator is a projector. Hence, an operator $\hat{E}$ is a projector if $\hat{E}^2 = \hat{E}$. The projection is along its null space and onto its range. When these subspaces are orthogonal $\hat{E}$ is called an orthogonal projector, which is the case if and only if $\hat{E}$ is self-adjoint. Otherwise it is called *oblique projector*.

Given two closed subspaces, $\mathcal{V} \in \mathcal{H}$ and $\mathcal{W}^\perp \in \mathcal{H}$, such that $\mathcal{S} = \mathcal{V} + \mathcal{W}^\perp$ and $\mathcal{V} \cap \mathcal{W}^\perp = \{0\}$, the oblique projector operator onto $\mathcal{V}$ along $\mathcal{W}^\perp$ will be represented as $\hat{E}_{\mathcal{VW}^\perp}$. Then $\hat{E}_{\mathcal{VW}^\perp}$ satisfies $\hat{E}^2_{\mathcal{VW}^\perp} = \hat{E}_{\mathcal{VW}^\perp}$ and, consequently,

$$
\begin{aligned}
\hat{E}_{\mathcal{VW}^\perp} f &= f, \quad \text{if} \quad f \in \mathcal{V} \\
\hat{E}_{\mathcal{VW}^\perp} f &= 0, \quad \text{if} \quad f \in \mathcal{W}^\perp.
\end{aligned}
$$

In the particular case for which $\mathcal{W}^\perp = \mathcal{V}^\perp$ the operator $\hat{E}_{\mathcal{VV}^\perp}$ is an orthogonal projection onto $\mathcal{V}$. For indicating an orthogonal projector onto a subspace, $\mathcal{X}$ say, we use the particular notation $\hat{P}_\mathcal{X}$.

Consider that $\{v_i\}_{i=1}^M$ is a spanning set for $\mathcal{V}$ and $\{u_i\}_{i=1}^M$ a spanning set for $\mathcal{W}$, which is the orthogonal complement of $\mathcal{W}^\perp$ in $\mathcal{S}$, i.e., $\mathcal{S} = \mathcal{W} \oplus \mathcal{W}^\perp$, with $\oplus$ indicating the orthogonal sum. Thus the spanning sets of $\mathcal{V}$ and $\mathcal{W}$ satisfy $\{u_i\}_{i=1}^M = \{\hat{P}_\mathcal{W} v_i\}_{i=1}^M$. If the condition $\mathcal{V} \cap \mathcal{W}^\perp = \{0\}$ is fulfilled, the operator $\hat{E}_{\mathcal{VW}^\perp}$ can be constructed as

$$
\hat{E}_{\mathcal{VW}^\perp} = \sum_{i=1}^M v_i \langle w_i, \cdot \rangle, \tag{2}
$$

where the operation $\langle w_i, \cdot \rangle$ indicates that $\hat{E}_{\mathcal{VW}^\perp}$ acts by performing inner products. The vectors $\{w_i\}_{i=1}^M$ in (2) are obtained from vectors $\{u_i\}_{i=1}^M$ through the equation [17,21]

$$
w_i = \sum_{j=1}^M g^\dagger_{i,j} u_j \tag{3}
$$

with $g^\dagger_{i,j}$ the element (i,j) of a matrix $G^\dagger$, a pseudo inverse of the matrix G the elements of which are given by the inner products $\langle u_i, v_j \rangle, i, j = 1, \ldots, M$.

Remark 1. *The pseudo inverse allows for the possibility of the spanning sets of $\mathcal{V}$ and $\mathcal{W}$ being redundant. However, the condition $\mathcal{V} \cap \mathcal{W}^\perp = \{0\}$ implies that $\mathcal{V}$ and $\mathcal{W}$ should have the same dimension and therefore the rank of G equals the dimension of $\mathcal{V}$ and $\mathcal{W}$.*

For later convenience, we introduce at this point an alternative representation of $\hat{E}_{\mathcal{VW}^\perp}$. To this end, denoting as $\boldsymbol{e}_i, i = 1, \ldots, M$ the standard orthonormal basis in $\mathbb{C}^M$, we define the operators $\hat{V} : \mathbb{C}^M \to \mathcal{V}$ and $\hat{W} : \mathbb{C}^M \to \mathcal{W}$ as

$$
\hat{V} = \sum_{i=1}^M v_i \langle \boldsymbol{e}_i, \cdot \rangle, \qquad \hat{W} = \sum_{i=1}^M u_i \langle \boldsymbol{e}_i, \cdot \rangle.
$$

Thus the corresponding adjoint operators $\hat{W}^* : \mathcal{W} \to \mathbb{C}^M$ and $\hat{V}^* : \mathcal{V} \to \mathbb{C}^M$ are

$$
\hat{V}^* = \sum_{i=1}^M \boldsymbol{e}_i \langle v_i, \cdot \rangle, \qquad \hat{W}^* = \sum_{i=1}^M \boldsymbol{e}_i \langle u_i, \cdot \rangle.
$$

It follows from the above definitions that $\hat{P}_\mathcal{W} \hat{V} = \hat{W}$ and $\hat{W}^* \hat{P}_\mathcal{W} = \hat{W}^*$.

Considering that $\psi_n \in \mathbb{C}^M$, $n = 1, \ldots, M$, are the eigenvectors of matrix $G = \hat{W}^* \hat{V} = \hat{W}^* \hat{W}$, and assuming that there exist N nonzero eigenvalues λ_n, $n = 1, \ldots, N$, on ordering these eigenvalues in descending order we can express the matrix elements of the Moore-Penrose pseudo inverse of G as:

$$g_{i,j}^\dagger = \sum_{n=1}^{N} \psi_n(i) \frac{1}{\lambda_n} \psi_n^*(j), \tag{4}$$

with $\psi_n(i)$ the i-th component of ψ_n. Moreover, the orthonormal vectors

$$\xi_n = \frac{\hat{W}\psi_n}{\sigma_n}, \quad \sigma_n = \sqrt{\lambda_n}, \quad n = 1, \ldots, N \tag{5}$$

are singular vectors of $\hat{W}$, which satisfies $\hat{W}^* \xi_n = \sigma_n \psi_n$, as it is immediate to verify. By defining now the vectors η_n, $n = 1, \ldots, N$ as

$$\eta_n = \frac{\hat{V}\psi_n}{\sigma_n}, \quad n = 1, \ldots, N, \tag{6}$$

the projector $\hat{E}_{\mathcal{V}\mathcal{W}^\perp}$ in (2) is recast in the fashion

$$\hat{E}_{\mathcal{V}\mathcal{W}^\perp} = \sum_{n=1}^{N} \eta_n \langle \xi_n, \cdot \rangle. \tag{7}$$

Inversely, the representation (2) of $\hat{E}_{\mathcal{V}\mathcal{W}^\perp}$ arises from (7), since

$$w_i = \sum_{n=1}^{N} \xi_n \frac{1}{\sigma_n} \psi_n^*(i), \quad i = 1, \ldots, M. \tag{8}$$

Proposition 1. *The vectors $\xi_n \in \mathcal{W}$, $n = 1, \ldots, N$ and $\eta_n \in \mathcal{V}$, $n = 1, \ldots, N$ given in (5) and (6) are biorthogonal to each other and span $\mathcal{W}$ and $\mathcal{V}$, respectively.*

Proof. Using (5) and (6) we have

$$\langle \xi_m, \eta_n \rangle = \frac{1}{\sigma_n \sigma_m} \langle \hat{W}\psi_n, \hat{V}\psi_m \rangle = \frac{1}{\sigma_n \sigma_m} \langle \psi_n, \hat{W}^* \hat{V}\psi_m \rangle = \delta_{n,m} \frac{\lambda_m}{\sigma_n \sigma_m} = \delta_{n,m}, \tag{9}$$

which proves the biorthogonality property.

The proof that $\text{span}\{\xi_n\}_{n=1}^N = \mathcal{W}$ stems from the fact that $\mathcal{W} = \text{span}\{u_i\}_{i=1}^M = \text{span}\{w_i\}_{i=1}^M$, which allows us to express an arbitrary $g \in \mathcal{W}$ as the linear combination $g = \sum_{i=1}^M a_i w_i$. Then, using (8), we have $g = \sum_{n=1}^N \tilde{a}_n \xi_n$ with $\tilde{a}_n = \frac{1}{\sigma_n} \sum_{i=1}^M a_i \psi_n^*(i)$, which proves that $\mathcal{W} \subset \text{span}\{\xi_i\}_{i=1}^N$. On the other hand for $g \in \text{span}\{\xi_i\}_{i=1}^N$ we can write $g = \sum_{n=1}^N d_n \xi_n$ and using (5) we have $f = \sum_{i=1}^M \tilde{d}_i u_i$, with $\tilde{d}_i = \frac{1}{\sigma_n} \sum_{n=1}^N d_n \psi_n(i)$. This proves that $\text{span}\{\xi_i\}_{i=1}^N \subset \mathcal{W}$ and therefore $\text{span}\{\xi_n\}_{n=1}^N = \mathcal{W}$. The proof that $\text{span}\{\eta_n\}_{n=1}^N = \mathcal{V}$ is equivalent to the previous one. $\square$

Since $\hat{E}_{\mathcal{V}\mathcal{W}^\perp} f = f$ for every signal $f \in \mathcal{V}$, regardless of the subspace $\mathcal{W}^\perp$, one can consider a different subspace $\mathcal{W}^\perp$ to construct measurement vectors and reconstruct a signal in $\mathcal{V}$ using any set of such vectors, as long as $\mathcal{V} \cap \mathcal{W}^\perp = \{0\}$ with $\mathcal{W}^\perp$ the orthogonal complement of $\mathcal{W} = \text{span}\{w_i\}_{i=1}^M$. On the other hand, if $f \notin \mathcal{V}$, the measurement vectors can be chosen to be suitable for the particular processing task. For instance, if the goal is to produce an approximation $f_\mathcal{V}$ of f in the subspace $\mathcal{V}$, then in oder to minimize

the distance $||f - f_{\mathcal{V}}||$ we need $f_{\mathcal{V}} = \hat{P}_{\mathcal{V}}f$. Any other projection would yield a distance $||f - \hat{E}_{\mathcal{V}\mathcal{W}^\perp}f||$ which satisfies [16]

$$||f - \hat{P}_{\mathcal{V}}f|| \leq ||f - \hat{E}_{\mathcal{V}\mathcal{W}^\perp}f|| \leq \frac{1}{\cos(\theta)}||f - \hat{P}_{\mathcal{V}}f||,$$

where θ is the minimum angle between the subspaces $\mathcal{V}$ and $\mathcal{W}$. The equality is attained for $\mathcal{V} = \mathcal{W}$, which correspond to the orthogonal projection.

However, if the aim were to discriminate from a signal produced by different phenomena only the component in $\mathcal{V}$, then, as discussed below, an oblique projection turns to be appropriate.

Suppose that a signal f is the superposition of two signals f_1 and f_2 with $f_1 \in \mathcal{V}$ and f_2 in $\mathcal{W}^\perp$. The projection that will rescue f_1 from f is $\hat{E}_{\mathcal{V}\mathcal{W}^\perp}f$. A number of signal processing examples where an oblique projection is required are given in [1]. Provided that the subspaces hosting the signal components are well separated, the discrimination of components with different structure is successful. Unfortunately, this is not always the case and the construction of the necessary projector may generate an ill posed problem.

3   The need for sparse representations in the present context

This section is dedicated to illustrate, by recourse to a numerical example, the crucial role that the search for sparse solutions plays in the construction of oblique projectors for signal discrimination. Consider that the spaces $\mathcal{V}$ and $\mathcal{W}^\perp$, such that $\mathcal{V} \cap \mathcal{W}^\perp = \{0\}$, are given, and the spanning set for $\mathcal{V}$ is a basis of dimension M. For constructing the dual vectors w_i as in (8) we first construct $\hat{P}_{\mathcal{W}^\perp}$, to generate the vectors

$$u_i = v_i - \hat{P}_{\mathcal{W}^\perp}v_i, \quad i = 1, \ldots, M \tag{10}$$

spanning $\mathcal{W}$.

Remark 2. *Since $\mathcal{V} \cap \mathcal{W}^\perp = \{0\}$, and the set $\{v_i\}_{i=1}^M$ is assumed to be linearly independent, the set $\{u_i\}_{i=1}^M$ is also linearly independent.*

Proof. Suppose that $\sum_{i=1}^M b_i u_i = 0$ for some set of numbers $\{b_i\}_{i=1}^M$. Then (10) implies $g = \hat{P}_{\mathcal{W}^\perp}g$ for $g = \sum_{i=1}^M b_i v_i$. Since by definition $g \in \mathcal{V}$, and $\mathcal{V} \cap \mathcal{W}^\perp = \{0\}$ by hypothesis, we conclude that $g = 0$. Hence, the fact that the set $\{v_i\}_{i=1}^M$ is linearly independent implies that $b_i = 0$, $i = 1, \ldots, M$, which establishes that the set $\{u_i\}_{i=1}^M$ is linearly independent. ◻

Remark 3. *Conversely, the fact that nonzero vectors constructed as in (10) are linearly independent implies that $\mathcal{V} \cap \mathcal{W}^\perp = \{0\}$ [15].*

In order to render the numerical calculation of the vectors w_i, $i = 1, \ldots, M$ spanning $\mathcal{W}$ as stable as possible, it is convenient to orthonormalize vectors u_i, $i = 1, \ldots, M$ to obtain the vectors q_i, $i = 1, \ldots, M$ satisfying $\langle q_i, q_j \rangle = \delta_{i,j}$. With these vectors we construct the $M \times M$ matrix G having elements $\langle q_i, v_j \rangle$. For the situation considered here this matrix has an inverse. Let us denote the element (i, j) of G^{-1} as $g_{i,j}^{-1}$ and construct the corresponding vectors w_i, $i = 1, \ldots, M$ as prescribed in (3) or (8). As will be illustrated by the numerical example below, in spite of the fact that 'theoretically' $\mathcal{V} \cap \mathcal{W}^\perp = \{0\}$, numerical errors, due to the existence of small singular values, may cause the failure to find the unique signal splitting that theoretically one should expect.

Example 1. Let $\mathcal{V}$ be the cardinal cubic spline space with distance 0.065 between consecutive knots, on the interval $[0, 10]$. This is a subspace of dimension $M = 163$, which we span using a B-spline basis $\{B_i(x),\, x \in [0, 10]\}_{i=1}^{163}$. The background we wish to filter belongs to the subspace $\mathcal{W}^\perp$ spanned by the set of functions $\{y_i(x) = (x + 1)^{-0.05i},\, x \in [0, 10]\}_{i=1}^{50}$. Here the inner product is defined as $\langle f, g \rangle = \int_0^{10} f(x)^* g(x)\, dx$, and all the integrals are computed numerically.

This example is very illustrative of how sensitive to numerical errors the computation of oblique projectors is. The subspace we are dealing with are disjoint: the last five singular values of the corresponding matrix G are:

$$0.2305, 0.2298, 9.3211 \times 10^{-4}, 2.5829 \times 10^{-6}, 2.5673 \times 10^{-7},$$

while the first is $\sigma_1 = 1.5018$. The smallest singular value cannot be considered a numerical representation of zero when the calculations are being carried out in double precision arithmetic. Hence, one can assert that the condition $\mathcal{V} \cap \mathcal{W}^\perp = \{0\}$ is fulfilled. However, due to the three small singular values the computation of the measurement vectors in the whole subspace $\mathcal{W}$ is inaccurate enough to cause the failure to correctly separate signals in $\mathcal{V}$ from their background. The left graph of Figure 1 is generated by a random superposition or $K = 70$ B-splines added to a background in the given $\mathcal{W}^\perp$. The broken line in the right graph represents the oblique projection onto the given $\mathcal{V}$ along $\mathcal{W}^\perp$. As can be seen, the projection does not produce the required signal, which is represented by the continuous dark line in the same graph. Now, since the spectrum of singular values has a clear jump (the last three singular values are far from the previous ones) it might seem that one could regularize the calculation by truncation of singular values. Nevertheless, such a methodology turns out to be not appropriate for the present problem, as it does not yield the correct separation. The light lines in the right graph of Figure 1 depict the three approximations obtained by neglecting one, two and three singular values.

Propositions 2 below analyzes the effect that regularization by truncation of singular values produces in the resulting projection.
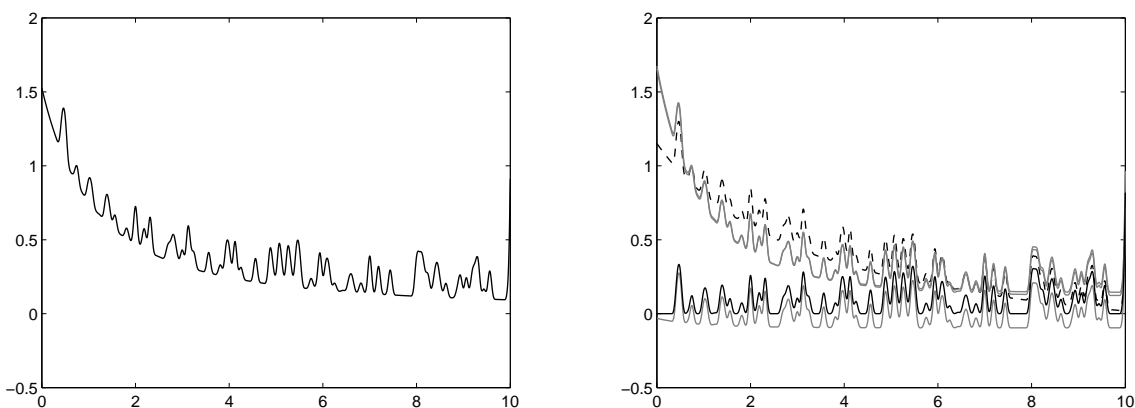

Figure 1: Left graph: signal plus background. Right graph: the dark continuous line corresponds to the signal to be discriminated from the one in the left graph. The broken line corresponds to the approximation resulting from the oblique projection. The light lines correspond to the approximations obtained by truncation of singular values (the one closest to the required signal correspond to the truncation of three singular values).

Proposition 2. *Truncation of the expansion* (7) *to consider up to r terms, produces an oblique projector along* $\tilde{\mathcal{W}}_r = \mathcal{W}^\perp + \tilde{\mathcal{W}}_0 + \tilde{\mathcal{V}}_0$, *with* $\tilde{\mathcal{V}}_0 = \text{span}\{\eta_i\}_{i=r+1}^N$ *and* $\tilde{\mathcal{W}}_0 = \text{span}\{\xi_i\}_{i=r+1}^N$, *onto* $\tilde{\mathcal{V}}_r = \text{span}\{\eta_i\}_{i=1}^r$.

Proof. The biorthogonality between $\{\xi\}_{i=1}^r$ and $\{\eta_i\}_{i=1}^r$ established in Proposition 1 ensures that $\hat{E}_{\tilde{\mathcal{V}}_r \tilde{\mathcal{W}}_r} = \sum_{i=1}^r \eta_i \langle \xi_i, \cdot \rangle$ is a projector, since $\hat{E}_{\tilde{\mathcal{V}}_r \tilde{\mathcal{W}}_r}^2 = \hat{E}_{\tilde{\mathcal{V}}_r \tilde{\mathcal{W}}_r}$.

As established in Proposition 1, $\mathcal{V} = \text{span}\{\eta_i\}_{i=1}^N$, and therefore every $f \in \mathcal{V}$ can be decomposed as $f = f_r + f_o$ with $f_r \in \text{span}\{\eta_i\}_{i=1}^r$ and $f_o \in \text{span}\{\eta_i\}_{i=r+1}^N$. Moreover, $\hat{E}_{\tilde{\mathcal{V}}_r \tilde{\mathcal{W}}_r} f = f_r, \hat{E}_{\tilde{\mathcal{V}}_r \tilde{\mathcal{W}}_r} f_r = f_r$, and $\hat{E}_{\tilde{\mathcal{V}}_r \tilde{\mathcal{W}}_r} f_o = 0$, which proves that the projection is onto $\tilde{\mathcal{V}}_r$ and $\tilde{\mathcal{V}}_0$ is included in the null space of $\hat{E}_{\tilde{\mathcal{V}}_r \tilde{\mathcal{W}}_r}$. Equivalently, for every $g_o \in \tilde{\mathcal{W}}_0 = \text{span}\{\xi_i\}_{i=r+1}^N$ we have $\hat{E}_{\tilde{\mathcal{V}}_r \tilde{\mathcal{W}}_r} g_o = 0$, because the set $\{\xi_i\}_{i=1}^N$ is orthonormal. Thus, $\tilde{\mathcal{W}}_0$ is included in the null space of $\hat{E}_{\tilde{\mathcal{V}}_r \tilde{\mathcal{W}}_r}$. $\qquad\square$

3.1 Getting ready for a greedy search of the sparse solution

We discuss here the properties that will be of assistance in the next section, where we will present our strategy for the search of the sparse representation achieving the desired signal discrimination. The goal is to avoid the computation of the measurement vectors in the whole subspace. Instead, we strive to find the subspace $\mathcal{V}_K \subset \mathcal{V}$, where the signal component one wants to discriminate from the noise is assumed to lie. We work under the hypothesis that the subspace $\mathcal{W}^\perp$ is given and fixed. Furthermore, $\mathcal{V} \cap \mathcal{W}^\perp = \{0\}$, which implies that there exists a unique solution for the signal splitting. The problem we need to address arises from the fact that, if the subspaces $\mathcal{V}$ and $\mathcal{W}^\perp$ are not well separated, the numerical calculation of the measurement vectors is not accurate (due to the numerical operations being carried out in finite precision arithmetic). As a consequence, the representation of the corresponding projector fails to produce the correct signals separation.

Assuming that we are able to accurately compute in finite precision arithmetic r measurement vectors, we could attempt to filter structured noise of a signal belonging to a subspace spanned by at most r of such vectors (i.e. the expansion of the signal in $\mathcal{V}$ should have at most r nonzero coefficients). However, even possessing this knowledge about the signal, the problem of finding the right subspace would be in general intractable: out of a set of cardinality M there exist $\binom{M}{r}$ possible subsets of cardinality r. An adaptive strategy for the subspace selection, given a signal, is advanced in [15]. Before revising and extending that strategy we need to recall two relevant properties of oblique projectors.

Property 1. *The oblique projector* $\hat{E}_{\mathcal{V} \mathcal{W}^\perp}$ *satisfies* $\hat{P}_\mathcal{W} \hat{E}_{\mathcal{V} \mathcal{W}^\perp} = \hat{P}_\mathcal{W}$.

Proof. It readily follows by applying $\hat{P}_\mathcal{W}$ on both sides of (2) or (7). Since $\langle u_i, v_j \rangle = \langle u_i, u_j \rangle$, one has

$$\hat{P}_\mathcal{W} \hat{E}_{\mathcal{V} \mathcal{W}^\perp} = \sum_{i=1}^M u_i \langle w_i, \cdot \rangle = \hat{P}_\mathcal{W}. \qquad (11)$$

Moreover, since $\hat{P}_\mathcal{W} \eta_i = \xi_i$, considering (7) we have

$$\hat{P}_\mathcal{W} \hat{E}_{\mathcal{V} \mathcal{W}^\perp} = \sum_{i=1}^N \xi_i \langle \xi_i, \cdot \rangle = \hat{P}_\mathcal{W}. \qquad (12)$$

$\qquad\square$

Property 2. *Given a signal f in $\mathcal{V} + \mathcal{W}^\perp = \mathcal{W} \oplus \mathcal{W}^\perp$, the only vector $g \in \mathcal{V}$ satisfying*

$$\hat{P}_\mathcal{W} f = \hat{P}_\mathcal{W} g \tag{13}$$

is $g = \hat{E}_{\mathcal{V}\mathcal{W}^\perp} f$.

Proof. If $g = \hat{E}_{\mathcal{V}\mathcal{W}^\perp} f$ (13) trivially follows from Property 1. Let us assume now that there exists $g \in \mathcal{V}$ such that (13) holds. Then $\hat{P}_\mathcal{W}(f - g) = 0$, i.e., $(f - g) \in \mathcal{W}^\perp$. Hence $\hat{E}_{\mathcal{V}\mathcal{W}^\perp}(f - g) = 0$ and, since $g \in \mathcal{V}$, this implies that $\hat{E}_{\mathcal{V}\mathcal{W}^\perp} f = g$. □

Let us suppose that $\mathcal{V}_k = \text{span}\{v_i\}_{i=1}^k$ is given and the spanning set is linearly independent. Assuming that $\mathcal{V}_k \cap \mathcal{W}^\perp = \{0\}$ we guarantee that the set of vectors $\{u_i\}_{i=1}^k$, with u_i given in (10) is also linearly independent. Therefore the dimension of $\mathcal{V}_k$ is equal to the dimension of $\mathcal{W}_k = \text{span}\{u_i\}_{i=1}^k = \text{span}\{w_i^k\}_{i=1}^k$. We use now a superscript k to indicate that the measurement vectors $\{w_i^k\}_{i=1}^k$ span $\mathcal{W}_k$. Hence these vectors give rise to the oblique projection of a signal f, onto $\mathcal{V}_k$ and along $\mathcal{W}^\perp$, as given by:

$$\hat{E}_{\mathcal{V}_k \mathcal{W}^\perp} f = \sum_{i=1}^k v_i \langle w_i^k, f \rangle = \sum_{i=1}^k c_i^k v_i. \tag{14}$$

It is clear from (14) that if the atoms in the atomic decomposition were to be changed (or some atoms were added to or deleted from the decomposition) the measurement vectors w_i^k, and consequently the coefficients c_i^k in (14), would need to be modified. The recursive equations below provide an effective way of implementing the task.

Forward/backward adapting of measurement vectors

Starting with $w_1^1 = \frac{u_1}{||u_1||^2}$, and u_1 as in (10), the measurement vectors $w_i^{k+1}, i = 1 \ldots, k+1$ can be recursively constructed from w_i^k, $i = 1 \ldots, k$ as follows [21]:

$$w_i^{k+1} = w_i^k - w_{k+1}^{k+1} \langle u_{k+1}, w_i^k \rangle, \quad i = 1, \ldots, k \tag{15}$$

$$w_{k+1}^{k+1} = \frac{\gamma_{k+1}}{||\gamma_{k+1}||^2}, \quad \gamma_{k+1} = u_{k+1} - \hat{P}_{\mathcal{W}_k} u_{k+1}, \tag{16}$$

where $\hat{P}_{\mathcal{W}_k}$ is the orthogonal projector onto $\mathcal{W}_k = \text{span}\{u_i\}_{i=1}^k$. We note that, since $\hat{P}_\mathcal{W} w_i^k = w_i^k$, (15) can also be written as

$$w_i^{k+1} = w_i^k - w_{k+1}^{k+1} \langle v_{k+1}, w_i^k \rangle, \quad i = 1, \ldots, k. \tag{17}$$

It follows from the above equations that when incorporating a linearly independent atom v_{k+1} in the atomic decomposition (14), the coefficients can be conveniently modified according to the recursive equations

$$c_{k+1}^{k+1} = \langle w_{k+1}^{k+1}, f \rangle, \tag{18}$$

$$c_i^{k+1} = \langle w_i^{k+1}, f \rangle = c_i^k - c_{k+1}^{k+1} \langle w_i^k, v_{k+1} \rangle, \quad i = 1, \ldots, k. \tag{19}$$

Conversely, considering that the atom, v_j say, is to be removed from the atomic decomposition (14), and denoting the corresponding subspaces $\mathcal{V}_{k \backslash j}$ and $\mathcal{W}_{k \backslash j}$, in order to span $\mathcal{W}_{k \backslash j}$ the measurement vectors $w_i^{k \backslash j}$, $i = 1, \ldots, k$ are modified according to the equation [21]

$$w_i^{k \backslash j} = w_i^k - \frac{w_j^k \langle w_j^k, w_i^k \rangle}{||w_j^k||^2}, \quad i = 1, \ldots, j-1, j+1, \ldots, k. \tag{20}$$

Consequently, the coefficients in (14) should be changed to

$$c_i^{k \backslash j} = c_i^k - \frac{c_j^k \langle w_i^k, w_j^k \rangle}{||w_j^k||^2}, \quad i = 1, \ldots, j-1, j+1, \ldots, k. \tag{21}$$

4 Adaptive pursuit strategy for subspace selection

Given a signal f, we aim at finding the subspace $\mathcal{V}_K \subset \mathcal{V}$ where the signal belongs. Let us stress once again that the problem arises from the impossibility of correctly computing the measurement vectors spanning the whole subspace $\mathcal{W}$. Otherwise the right subspace is determined simply by the indices corresponding to the atoms having nonzero coefficients in the full atomic decomposition (14).

Since $\mathcal{V}_{k+1} + \mathcal{W}^\perp = \mathcal{W}_{k+1} \oplus \mathcal{W}^\perp$ the forward selection criterion we propose is based on Property 2, which implies that if a given f satisfies $\hat{P}_{\mathcal{W}}f = \hat{P}_{\mathcal{W}_k}f$ it also satisfies $\hat{E}_{\mathcal{V}\mathcal{W}^\perp}f = \hat{E}_{\mathcal{V}_k\mathcal{W}^\perp}f$. Thus, by fixing $\hat{P}_{\mathcal{W}_k}$, at iteration $k+1$ we select the index ℓ_{k+1} such that $||\hat{P}_{\mathcal{W}}f - \hat{P}_{\mathcal{W}_{k+1}}f||^2$ is minimized.

Proposition 3. *Let us denote J to the set of indices $i = 1, \ldots, M$. Given $\mathcal{W}_k$, the index ℓ_{k+1} corresponding the atom $u_{\ell_{k+1}}$ in the set $\{u_i\}_{i \in J}$ for which $||\hat{P}_{\mathcal{W}}f - \hat{P}_{\mathcal{W}_{k+1}}f||^2$ is minimal is to be determined as*

$$\ell_{k+1} = \arg \max_{n \in J \setminus J_k} \frac{|\langle \gamma_n, f \rangle|}{||\gamma_n||}, \; \gamma_n \neq 0, \tag{22}$$

with γ_n given in (16), and J_k the set of indices that have been previously chosen to determine $\mathcal{W}_k$.

Proof. It readily follows since $\hat{P}_{\mathcal{W}_{k+1}}f = \hat{P}_{\mathcal{W}_k}f + \frac{\gamma_n \langle \gamma_n, f \rangle}{||\gamma_n||^2}$ and hence $||\hat{P}_{\mathcal{W}}f - \hat{P}_{\mathcal{W}_{k+1}}f||^2 = ||\hat{P}_{\mathcal{W}}f|| - ||\hat{P}_{\mathcal{W}_k}f||^2 - \frac{|\langle \gamma_n, f \rangle|^2}{||\gamma_n||^2}$. Because $\hat{P}_{\mathcal{W}}f$ and $\hat{P}_{\mathcal{W}_k}f$ are fixed, $||\hat{P}_{\mathcal{W}}f - \hat{P}_{\mathcal{W}_{k+1}}f||^2$ is minimized if $\frac{|\langle \gamma_n, f \rangle|}{||\gamma_n||}$, $\gamma_n \neq 0$ is maximal over all $n \in J \setminus J_k$. $\square$

Remark 4. *Since $\hat{P}_{\mathcal{W}_{k+1}} = \hat{P}_{\mathcal{W}}\hat{E}_{\mathcal{V}_{k+1}\mathcal{W}^\perp}$ we can write $||\hat{P}_{\mathcal{W}}f - \hat{P}_{\mathcal{W}_{k+1}}f||^2 = ||\hat{P}_{\mathcal{W}}(f - \hat{E}_{\mathcal{V}_{k+1}\mathcal{W}^\perp}f)||^2$, and the condition of the previous proposition can be seen as the condition for minimizing the distance of $\hat{E}_{\mathcal{V}_{k+1}\mathcal{W}^\perp}f$ to f, with respect to the weighted seminorm $|| \cdot ||_{P_{\mathcal{W}}}$ induced by the weighted inner product $\langle \cdot, \cdot \rangle_{P_{\mathcal{W}}}$ defined as $\langle f, g \rangle_{P_{\mathcal{W}}} = \langle f, P_{\mathcal{W}}g \rangle$.*

The OBMP selection criterion given in [15], which is based on the consistency principle [16,17], selects the index ℓ_{k+1} as the maximizer over $n \in J \setminus J_k$ of

$$\frac{|\langle \gamma_n, f \rangle|}{||\gamma_n||^2}, \quad ||\gamma_n|| \neq 0.$$

This condition was proposed in [15] so as to select the measurement vector w_{k+1}^{k+1} producing the maximum consistency error $\Delta = |\langle w_{k+1}^{k+1}, f - \hat{E}_{\mathcal{V}_k\mathcal{W}^\perp}f \rangle|$, with regard to a new measurement w_{k+1}^{k+1}. However, since the measurement vectors are not normalized to unity, it is sensible to consider the consistency error relative to the corresponding measurement vector norm $||w_{k+1}^{k+1}||$, and select the index so as to maximize over $k+1 \in J \setminus J_k$ the relative consistency error

$$\tilde{\Delta} = \frac{|\langle w_{k+1}^{k+1}, f - \hat{E}_{\mathcal{V}_k\mathcal{W}^\perp}f \rangle|}{||w_{k+1}^{k+1}||}, \quad ||w_{k+1}^{k+1}|| \neq 0 \tag{23}$$

Property 3. *The index ℓ_{k+1} satisfying (22) maximizes over $k+1 \in J \setminus J_k$ the relative consistency error (23)*

Proof. Since for all vector w_{k+1}^{k+1} given in (16) $\hat{E}_{\mathcal{V}_k \mathcal{W}^\perp} w_{k+1}^{k+1} = 0$ and $||w_{k+1}^{k+1}|| = ||\gamma_{k+1}||^{-1}$ we have

$$\tilde{\Delta} = \frac{|\langle w_{k+1}^{k+1}, f \rangle|}{||w_{k+1}^{k+1}||} = \frac{|\langle \gamma_{k+1}, f \rangle|}{||\gamma_{k+1}||}.$$

Hence, maximization of $\tilde{\Delta}$ over $k + 1 \in J \setminus J_k$ is equivalent to (22). □

It is clear at this point that the forward selection of indices prescribed by proposition (22) is equivalent to selecting the indices by applying the Optimized Orthogonal Matching Pursuit (OOMP) [22] strategy on the projected signal $\hat{P}_{\mathcal{W}} f$ using the dictionary $\{u_i\}_{i \in J}$.

The hypothesis that the computation of more than r measurement vectors becomes an ill posed problem enforces the forward selection of indices to stop if iteration r is reached. Nevertheless, the fact that the signal is assumed to be K-sparse, with $K \leq r$, does not imply that before (or at) iteration r one will always find the correct subspace. The r-value just indicates that it is not possible to continue with the forward selection, because the computations would become inaccurate and unstable. Hence, if the right solution was not yet found, one needs to implement a strategy accounting for the fact that it is not feasible to compute more than r measurement vectors. An adequate procedure is achieved by means of the swapping-based refinement to the OOMP approach introduced in [23]. As discussed below, it consists of interchanging already selected atoms with nonselected ones.

Consider that at iteration r the correct subspace has not appeared yet and the selected indices are labeled by the r indices $\ell_1, \ldots, \ell_r$. In order to choose the label of the atom that minimizes the norm of the residual error as passing from approximation $\hat{P}_{\mathcal{W}_r} f$ to approximation $\hat{P}_{\mathcal{W}_{r \setminus j}} f$ we should fix the index of the atom to be deleted, ℓ_j say, as the one for which the quantity

$$\frac{|c_i^r|}{||w_i^r||} \tag{24}$$

is minimized $i = 1, \ldots, r$ [23, 24].

The process of eliminating one atom from the atomic decomposition (14) is called *backward step* while the process of adding one atom is called *forward step*. The forward selection criterion to choose the atom to replace the one eliminated in the previous step is accomplished by finding the index $\ell_i, i = 1, \ldots, r$ for which the the functional

$$e_n = \frac{|\langle \nu_n, f \rangle|}{||\nu_n||}, \quad \text{with} \quad \nu_n = u_n - \hat{P}_{\mathcal{W}_{r \setminus j}} u_n, \quad ||\nu_n|| \neq 0 \tag{25}$$

is maximized. In our framework, using (20), the projector $\hat{P}_{\mathcal{W}_{r \setminus j}}$ is computed as

$$\hat{P}_{\mathcal{W}_{r \setminus j}} = \hat{P}_{\mathcal{W}_r} - \frac{\langle w_i^r, w_j^r \rangle \langle w_j^r, \cdot \rangle}{||w_j^r||^2}.$$

Since $\hat{P}_{\mathcal{W}_r}$ and w_j^r are available, the computation of the sequence ν_n in (25) is a simple operation.

As proposed in [23] the swapping of pairs of atoms is repeated until the swapping operation, if carried out, would not decrease the approximation error. The implementation details for an effective realization of this process are given in [23], and MATLAB codes are available at [25]. Since there is no guarantee that at the end of the swapping of pairs of atoms the correct subspace has been found, the process can continue by increasing the number of atoms the swapping involves. At the second stage, in line with [26] we propose the swapping to be realized by the combinations of two backward steps followed by two

10

forward steps, provided that the interchange of the two atoms improves the approximation error. If at the end of the second stage the right subspace has not yet been found, the number of atoms involved in the swapping is increased up to three and so on. Notice that if the number of atoms to be interchanged reaches the value r the whole process would repeat identically. This is avoided by initiating the new circle with a different initial atom. Although convergence cannot be guaranteed, the above specified hypothesis ensure that the algorithm will stop when the correct signal splitting has been found. At such a stage one has $\hat{P}_\mathcal{W} f = \hat{P}_{\mathcal{W}_r} f$ with $\mathcal{W}_r$ spanned by the selected atoms $\ell_1, \ldots, \ell_r$. If the order K of sparseness of the signal is less than r a number of $r - K$ coefficients in the atomic decomposition

$$f = \sum_{i=1}^{r} v_{\ell_i} \langle w_i^r, f \rangle = \sum_{i=1}^{r} c_i^r v_{\ell_i}$$

will have zero value.

4.1 Examples

Firstly we applied the proposed strategy to the numerical simulation of Example 1, which is a very simple test for our method and therefore in a run of 50 simulations we could produce the correct signals splitting at the stage involving forward selection only.

Example 2. For this example we have used the same background as in Example 1, but a dictionary of B-splines spanning the same space as the basis. The dictionary consists of functions of broader support than the basis functions for the same space, and the translation parameter is reduced (for more details on the construction of B-splines dictionaries see [27], MATLAB codes are available at [25]). In this case, the spectrum of singular values of matrix G decreases continuously, as shown in the top graphs of Figure 2. Since it is difficult to decide on where to truncate the singular values, for the sake of comparison with the proposed technique we made a signal dependent truncation. This was achieved by setting the number Q of singular values to be considered so as to minimize

$$||\hat{P}_\mathcal{W} f - \hat{P}_{\tilde{\mathcal{W}}_Q} f||,$$

where $\tilde{\mathcal{W}}_Q$ indicates the subspace spanned by the first Q singular vectors of the operator $\hat{W}$ (c.f. (5)) Neither in this case the regularization by truncation of singular values was successful. The result is depicted by the lighter line in the right bottom graph of Figure 2. The dark line plots the sought signal.

By means of the proposed strategy we were able to find the correct signal splitting in the 50 simulations we ran. Only in one of the cases a re-initialization took place. Increasing the number of atoms in the simulated atomic decomposition up to 80, we also found the correct splitting in all the cases. In this simulation the re-initialization stage occurred in the case of 5 signals, out of the 50 signal in the run. By increasing the number of atoms up to 90, re-initialization took place in the case of 8 signals. As could be expected, due to the singular values decay, by increasing the number of atoms up to 100 in the atomic decomposition we started to observe instability in the calculations.

Example 3. Here the signal space is spanned by $M = 210$ vectors in R^L, with $L = 1000$, given as

$$\{ v_i = \cos(\frac{\pi(2j - 1)(i - 1)}{2L}), \quad j = 1, \ldots, L \}_{i=1}^{M}.$$
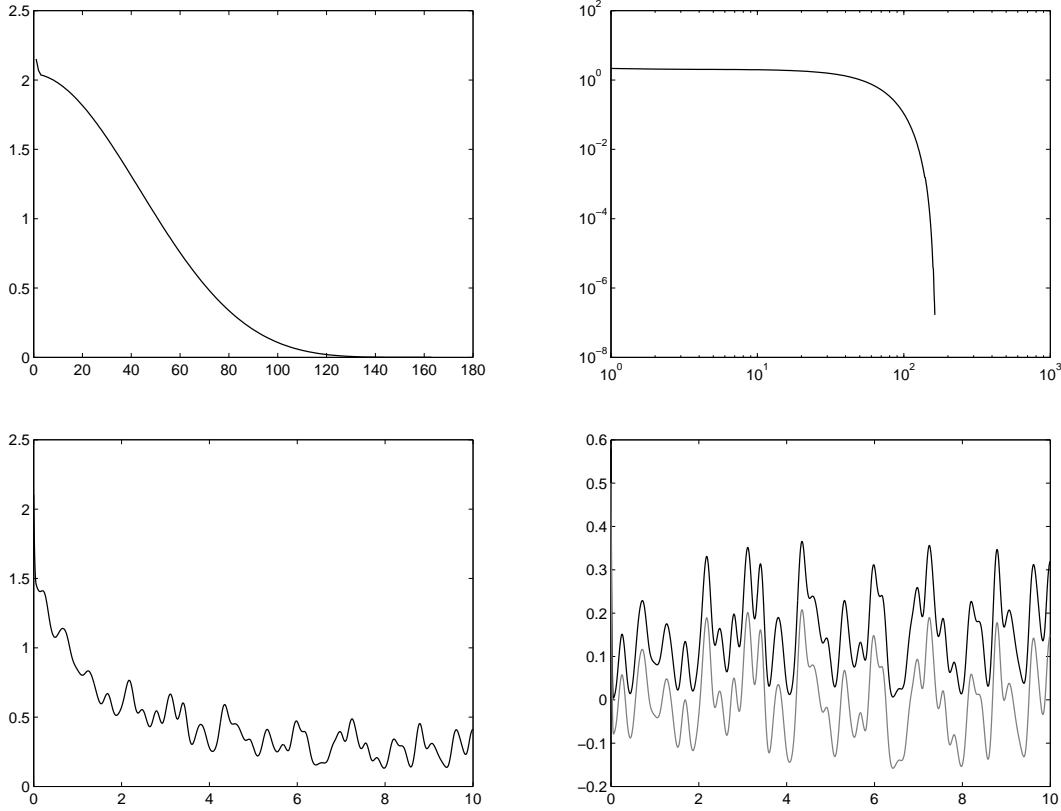
11

Figure 2: Top graphs: spectrum of singular values of matrix G in Example 2. The right graph plots the line in the left graph, but in logarithmic scale.

Bottom graphs: The one on the left shows the signal plus background. The dark line in the right one is the signal component to be discriminated from the signal in the left graph. The light line represents the approximation obtained by truncation of singular values. The proposed approach reproduces exactly the dark line.

The space of the noise is spanned by the set

$$\{y_i = e^{-35000(j-0.005i)^2}, \quad j = 1, \ldots, L\}_{i=1}^{400}.$$

We ran 50 simulations, keeping the noise fixed and considering a different realization of the signal, which was generated as a linear combination of 90 vectors taken randomly from the given spanning set.

The spectrum of singular values is depicted in the top graphs of Figure 3. In this case the signal dependent criterion for truncation does achieve the correct signal splitting. The left bottom graph of Figure 3 shows one of the realizations of the signal plus noise in the simulation. The dark line in the right graph of Figure 3 plots the exact signal. The approximation obtained by truncation of singular values is plotted with a lighter line, which cannot be distinguished from the dark one in the scale of the figure. It is clear from this result that in this case the signal does not have a significant component in the subspace spanned by the neglected singular vectors. Similar results are obtained in all the other realizations in the simulation. The norm of the error in this case is 0.53, while the mean value of the error norm with respect to the 50 cases is 0.78.

By applying the proposed strategy for searching the sparse representation we found the exact solution in the 50 cases. Re-initialization was necessary in 5 of the 50 realizations.
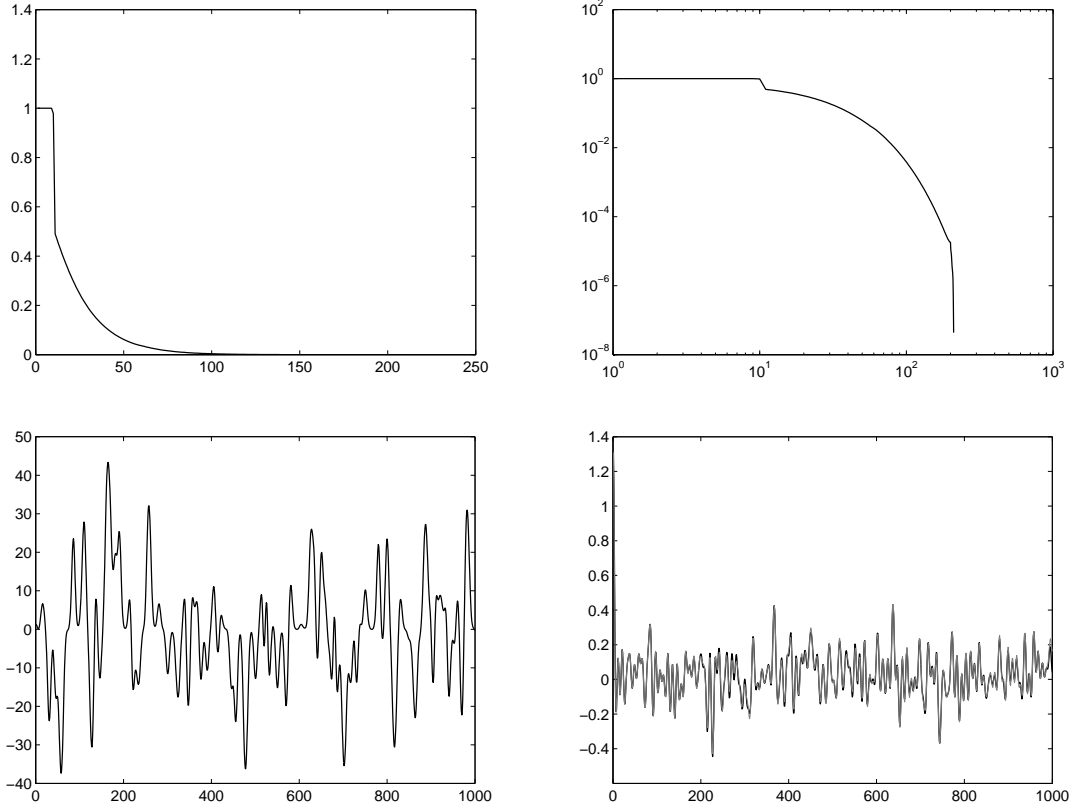

Figure 3: Top graphs: spectrum of singular values of matrix G in Example 3. The right graph plots the line in the left graph, but in logarithmic scale.
Bottom graphs: The one on the left shows the signal plus noise. The dark line in the right one is the signal to be separated from the one in the left graph. The approximation obtained by truncation of singular values is plotted by a lighter line, which cannot be distinguished from the other in the scale of the figure. The proposed approach reproduces the exact signal.

5   Conclusion

The role of sparse representations in the context of structured noise filtering has been discussed. The discrimination of signal components is achieved by an oblique projection onto the right subspace. Considerations were restricted to those cases for which the signal subspace and the noise subspace are theoretically complementary, but the construction of the dual basis for the whole signal subspace yields an ill posed problem, due to the calculations being carried out in finite precision arithmetic. It was shown by numerical simulations that, if the signal is sparse in a spanning set for the signal subspace, the required signal splitting may be achieved by means of adaptive techniques capable of searching for the required subspace while maintaining stability in the calculations. Although convergence of the proposed strategy for adaptive subspace search is not guaranteed, the method is capable to stop when the correct signal splitting is accomplished.

For the sake of comparison, an alternative regularization technique based on adaptive truncation of singular values was analyzed. The main disadvantage of such a technique lies in the fact that regularization is performed by a change of subspaces. Consequently, in general, the technique does not produce the required signal splitting. Moreover, even when a satisfactory splitting is attained (c.f. Example 3) the method does not provide an indication that this is so. Except for the very particular case in which the signal at hand has zero projection onto the subspace spanned by the disregarded singular vectors, the exact solution cannot be produced by this technique. On the contrary, the approach based on the search for the sparse representation is capable of producing the exact solution when the method stops. Thus, if the algorithm has converged, one can assert that the signal splitting is the required one.

References

[1] R. Behrens, L. Scharf, Signal processing applications of oblique projection operators, IEEE Transactions on Signal Processing 42 (1994) 1413–1424.

[2] R. DeVore, V. Temlyakov, Some remarks on greedy algorithms, Advances in Computational Mathematics 5 (1996) 173–187.

[3] G. Davis, S. Mallat, M. Avellaneda, Adaptive greedy approximations, Constructive Approximation 13 (1) (1997) 57–98.

[4] R. DeVore, Nonlinear approximation, Acta Numerica 7 (1998) 51–150.

[5] S. Chen, D. Donoho, M. Saunders, Atomic decomposition by basis pursuit, SIAM Journal on Scientific Computing 20 (1998) 33–61.

[6] V. Temlyakov, Greedy algorithms and M-term approximation with regard to redundant dictionaries, Journal of Approximation Theory 98 (1) (1999) 117–145.

[7] S. Aase, J. Husøy, K. Skretting, K. Engan, Optimized signal expansions for sparse representation, IEEE Trans. on Signal Process. 49 (5) (2001) 1087–1096.

[8] A. Gilbert, S. Muthukrishnan, M. Strauss, Approximation of functions over redundant dictionaries using coherence, in: the fourteenth annual ACM-SIAM symposium on Discrete algorithms, 2003, pp. 243 – 252.

[9] J. Tropp, Greed is good: algorithmic results for sparse approximation, IEEE Transactions on Information Theory 50 (10) (2004) 2231–2242.

[10] R. Gribonval, M. Nielsen, Nonlinear approximation with dictionaries. I. Direct estimates, Journal of Fourier Analysis and Applications 10 (2004) 55–71.

[11] D. Donoho, Compressed sensing, IEEE Trans. on Information Theory 52 (2006) 1289–1306.

[12] E. Candès, J. Romberg, Quantitative robust uncertainty principles and optimally sparse decompositions, Foundations of Comput. Math. 6 (2006) 227–254.

[13] E. Candès, T. Tao, Near optimal signal recovery from random projections: Universal encoding strategies?, IEEE Trans. on Information Theory 52 (2006) 5406–5425.

[14] R. Gribonval, R. F. i Ventura, P. Vandergheynst, A simple test to check the optimality of a sparse signal approximation, Signal Processing 86 (2006) 496–510.

[15] L. Rebollo-Neira, Oblique matching pursuit, IEEE Signal Processing Letters 11.

[16] M. Unser, A. Aldroubi, A general sampling theory for nonideal acquisition devices, IEEE Trans. Signal Processing 42 (1994) 2915–2925.

[17] Y. Eldar, Sampling with arbitrary sampling and reconstruction spaces and oblique dual frame vectors, Journal of Fourier Analysis and Applications 9 (2003) 77–96.

[18] M. Reed, B. Simon, Functional Analysis, Academic Press, London, 1980.

[19] M. Unser, J. Zerubia, Generalized sampling: Stability and performance analysis, IEEE Transactions on Signal Processing 45 (1997) 2941–2950.

[20] A. Hirabayashi, M. Unser, Consistent sampling and signal recovery, IEEE Transactions on Signal Processing 55 (2007) 4104–4115.

[21] L. Rebollo-Neira, Constructive updating/downdating of oblique projectors: a generalization of the gramschmidt process, Journal of Physics A: Mathematical and Theoritcal 40 (2007) 6381–6394.

[22] L. Rebollo-Neira, D. Lowe, Optimized orthogonal matching pursuit approach, IEEE Signal Processing Letters 9 (2002) 137–140.

[23] M. Andrle, L. Rebollo-Neira, A swapping-based refinement of orthogonal matching pursuit strategies, Signal Processing 86 (2006) 480–495.

[24] M. Andrle, L. Rebollo-Neira, E. Sagianos, Backward-optimized orthogonal matching pursuit approach, IEEE Signal Proc. Let. 11 (2004) 705–708.

[25] M. Andrle, L. Rebollo-Neira, Biorthogonal techniques for optimal signal representation, http://www.ncrg.aston.ac.uk/Projects/BiOrthog.

[26] M. Andrle, L. Rebollo-Neira, Improvement of orthogonal matching pursuit strategies by backward and forward movements, in: Proc. of the 31st International Conference on Acoustics, Speech, and Signal Processing, Vol. 3, 2006, pp. III–III.

[27] M. Andrle, L. Rebollo-Neira, Cardinal B-spline dictionaries on a compact interval, Appl. Comput. Harmon. Anal. 18 (2005) 336–346.